CHARONA

OCT : 7.200°

OFFICE OF PETITIONS



Max-Delbrück-Centrum für Molekulare Medizin

J. Behrens, W. Birchmeier

Mittel zur Diagnose und zur Therapie von Tumorerkrankungen

Zusammenfassung

Die Erfindung betrifft neue Wege zur Bekämpfung von Tumorerkrankungen durch Ausnutzung molekularbiologischer Zusammenhänge bei der Tumorentstehung.

Der Erfindung liegt die Aufgabe zugrunde, ein Verfahren zur Kontrolle der Regulation von ß-Catenin in Körperzellen zu entwickeln.

Gegenstand der Erfindung ist ein neues Protein, welches an ß-Catenin bindet und zu dessem zytoplasmatischen Abbau führt. Dieses Protein hat die Aminosäuresequenz gemäß Abb. 1 und wurde als CONDUCTIN bezeichnet.

Vom Vorkommen und der Wirkung des Conductins in Körperzellen abgeleitet werden Mittel zur Diagnose und zur Therapie von Tumorerkrankungen entwickelt.

74

Patentansprüche

- 1. Mittel zur Diagnose von Tumoren, enthaltend eine Substanz, mit der
- Conductin, seine Mutanten und Varianten oder Teile davon bzw.
- Gene, die für Conductin, seine Mutanten und Varianten oder Teile davon kodieren, bzw.
- m-RNA-Sequenzen, die von diesen Genen abgelesen werden, nachgewiesen werden.
- 2. Mittel zur Diagnose von Tumoren nach Anspruch 1, enthaltend spezifische Antikörper gegen Conductin, seine Varianten oder Mutanten oder Teile davon.
- 3. Mittel zur Diagnose von Tumoren nach Anspruch 1 und 2, dadurch gekennzeichnet, daß die spezifischen Antikörper monoklonale Antikörper sind.
- 4. Mittel zur Diagnose von Tumoren nach Anspruch 1, enthaltend korrespondierende Oligonukleotid-Primer bzw. DNA-Sonden zum Nachweis der Gene und deren Mutationen.
- 5. Mittel zur Diagnose von Tumoren nach Anspruch 1, enthaltend korrespondierende Oligonukleotid-Primer bzw. DNA-Sonden zum Nachweis der RNA-Sequenzen.
- 6. Mittel zur Therapie von Tumoren, enthaltend eine Substanz, die die Wirkung des Conductins im Körper aktiviert/reaktiviert.
- 7. Mittel nach Anspruch 6, enthaltend eine Substanz, die den Genpromoter des Conductins aktiviert.
- 8. Mittel nach Anspruch 6, enthaltend eine Substanz, die die Stabilität der mRNA-Sequenzen erhöht.
- 9. Mittel nach Anspruch 6, enthaltend eine Substanz, die die

75%

Aktivität des Conductins erhöht.

- 10. Conductin, seine Varianten und Mutanten sowie Teile davon.
- 11. Conductin nach Anspruch 10, gekennzeichnet durch die Aminosäuresequenz 1-839 gemäß Abb. 1, wobei Abb. 1 Bestandteil dieses Anspruchs ist.
- 12. Teilsequenz des Conductins nach Anspruch 10, gekennzeichnet durch die Aminosäuresequenz 78-200 (RGS-Domäne) der Abb. 1.
- 13. Teilsequenz des Conductins nach Anspruch 10, gekennzeichnet durch die Aminosäuresequenz 343-464 (B-Catenin-Bindungsdomäne)der Abb. 1.
- 14. Teilsequenz des Conductins nach Anspruch 10, gekennzeichnet durch die Aminosäuresequenz 782-832(Dishevelled Homologie-Region)der Abb. 1.
- 15. Teilsequenzen des Adenomatosis Poliposis Coli (APC), gekennzeichnet durch die Aminosäuresequenzen 1464-1604, 1516-1595, 1690-1778 und 1995-2083 als RGS-Domänen-Interaktionsorte.
- 16. cDNA-Sequenz von Conductin, seiner Varianten oder Mutanten oder Teilen davon.
- 17. cDNA-Sequenz des Conductins der Nukleotidfolge 1-3197 der Abb. 2, wobei Abb. 2 Bestandteil dieses Anspruchs ist.
- 18. cDNA-Teilsequenz des Conductins der Nukleotidfolge 452-820 (RGS-Genabschnitt)der Abb. 2.
- 19. cDNA-Teilsequenz des Conductins der Nukleotidfolge 1247-1612 (Genabschnitt der ß-Catenin-Bindungsdomäne)der Abb. 2.
- 20. cDNA-Teilsequenz des Conductins der Nukleotidfolge 2564-2716 (Genabschnitt der Dishevelled Homologie-Region)der Abb. 2.

21. Verwendung des Conductin-Gens für die Gentherapie von Tumor rkrankungen, dadurch gekennzeichnet, daß ein Vektor mit dem Conductin-Gen konstruiert wird, anschließend ein Gentransfer in den menschlichen Körper erfolgt und damit die Aktivität des Conductins in Körperzellen wiederhergestellt wird.

Beschreibung

F

Die Erfindung betrifft neue Wege zur Bekämpfung von Tumorerkrankungen durch Ausnutzung molekularbiologischer Zusammenhänge bei der Tumorentstehung. Sie betrifft im einzelnen ein Mittel zur Diagnose von Tumorerkrankungen, und darauf aufbauend ein Mittel zur Therapie. Sie betrifft ferner das neue Protein Conductin, seine Mutanten und Varianten sowie Teile davon, die dazu analogen cDNA-Sequenzen und deren Verwendung in gentherapeutischen und pharmakologischen Verfahren.

Anwendungsgebiete der Erfindung sind die Medizin und die pharmazeutische Industrie.

Cadherine und Catenine bilden Zelladhäsionskomplexe, die in zahlreichen Geweben für die Anheftung der Zellen aneinander verantwortlich sind. Die Cadherine sind Transmembranproteine und stellen den direkten Kontakt zwischen benachbarten Zellen her. α -, β - und γ -Catenin sind zytoplasmatische Komponenten, die die Cadherine mit dem Aktin-Zytoskelett verbinden. Neben Funktion bei der Zelladhäsion haben Catenine auch eine entscheidende Rolle bei Signaltransduktionsprozessen. B-Catenin in Vertebraten und das homologe Segmentpolaritäts-Genprodukt Armadillo in Drosphila werden durch den Wnt/Wingless-Signalweg stabilisiert (Nusse, R., Cell 89, 321-323, 1997). Dies führt zu einer Erhöhung der zytoplasmatischen, nicht an Cadherin gebundenen Fraktion dieser Proteine, die daraufhin mit HMG-Transkriptionsfaktoren der LEF-1/TCF-Familie wechselwirken können. Als Resultat wird B-Catenin/Armadillo in den Zellkern transportiert, wo es zusammen mit den LEF/TCF-Proteinen an DNA bindet und bestimmte Gene aktiviert (Behrens, J. et. al., Nature 382, 638-642, 1997).

Dieser Signalweg spielt auch eine Rolle bei der Tumorentstehung. In Kolonepithelzellen wird der zytoplasmatische Pool von ß-Catenin durch das Tumorsuppressor-Genprodukt APC (Adenomatosis Polyposis Coli) streng reguliert. Mutationen von APC, wie sie in

etwa 80% aller Kolonkarzinome auftreten, führen zu verkürzten Formen des APC Proteins, die nicht mehr in der Lage sind ß-Catenin zu destabilisieren. Dadurch findet man in diesen Tumoren permanente Komplexe von ß-Catenin mit dem HMG-Transkriptionsfaktor TCF-4, welche für die Transformation der Zellen verantwortlich gemacht werden. Diese Theorie wird gestützt durch den kürzlichen Befund, daß in Tumoren, in denen APC nicht verändert ist, Mutationen von ß-Catenin auftreten. Diese führen ebenfalls zur zytoplasmatischen Stabilisierung von ß-Catenin und zur Assoziation mit LEF-1/TCF-Faktoren (Morin, P.J. et. al., Science 275, 1787-1790).

Die Erfindung hat das Ziel, einen neuen Weg zur Verhinderung der Tumorentstehung zu finden. Der Erfindung liegt die Aufgabe zugrunde, ein Verfahren zur Kontrolle der Regulation von B-Catenin in Körperzellen zu entwickeln.

Gegenstand der Erfindung ist ein neues Protein, welches an ß-Catenin bindet und zu dessem zytoplasmatischen Abbau führt. Dieses Protein hat die Aminosäuresequenz gemäß Abb. 1 und wurde als CONDUCTIN bezeichnet.

Das Erfindung beruht nun auf der eigenen Erkenntnis, daß Conductin über eine ß-Catenin-Bindungsdomäne an ß-Catenin und über eine sogenannte RGS-Domäne (Regulator of G-Protein Signalling) an APC-Fragmente bindet. Dadurch kommt es zum zytoplasmatischen Abbau von ß-Catenin und in Vertebraten zur Blockade des Wnt/Wingless-Signalwegs. Damit ist klar, daß Conductin ein wichtiger Regulator der ß-Catenin-Funktion ist und im Zusammenspiel mit APC zur Tumorsuppression beiträgt.

Davon abgeleitet betrifft die Erfindung ein Mittel zur Diagnose von Tumorerkrankungen, welches dadurch gekennzeichnet ist, daß das Vorhandensein und die Menge von Conductin, seiner Mutanten und Varianten oder s iner Teile in Körperz llen nachgewiesen wird. Dieser Nachweis kann auf der Proteineben mit spezifischen Antikörpern durchgeführt werden, speziell mit monoklonalen

Antikörpern.

Die Diagnose von Tumorerkrankungen kann gemäß der Erfindung auch auf der Genebene erfolgen. Dazu werden mit ausgewählten Primern und cDNA-Sonden, die aus der Gensequenz des Conductins abgeleitet sind,

- das Gen, das für Conductin, seine Mutanten und Varianten oder Teile davon kodiert, bzw.
- mRNA-Sequenzen, die von diesen Genen abgelesen werden, nachgewiesen.

Das erfindungsgemäße Mittel zur Therapie von Tumorerkrankungen enthält Substanzen, die die Wirkung des Conductins im Körper aktivieren/reaktivieren. Das sind vor allem Mittel, die den Genpromoter des Conductins aktivieren bzw. Mittel, die die Stabilität der von den Conductin-Genen abgeleiteten m-RNA-Sequenzen erhöht. Das Hauptziel aller dieser Mittel besteht erfindungsgemäß darin, die Aktivität des Conductins in den Körperzellen zu erhöhen. Dazu kommen u. a. kleinmolekulare Substanzen in Betracht, die z. B. durch High-Througput-Number-Screening gefunden werden.

Die Erfindung umfaßt auch gentherapeutische Mittel, enthaltend Gene, die für Conductin, seine Mutanten und Varianten oder Teile davon kodieren, bzw. mRNA-Sequenzen, die von diesen Genen abgelesen werden.

Unter Schutz gestellt wird ferner das neue Protein Conductin gemäß Abb. 1, seine Mutanten und Varianten sowie Teile davon. Besonders bevorzugte Teilsequenzen sind die Aminosäuren 78-200 (RGS),343-464 (ß-Catenin-Bindungsdomäne) und 782-832 (Dishevelled Homologie-Region). Zum Schutzumfang gehören auch Teilsequenzen des Adenomatosis Poliposis Coli (APC), gekennzeichnet durch die Aminosäuresequenzen 1464-1604, 1516-1595, 1690-1778 und 1995-2083 als RGS-Domänen-Interaktionsorte.

Gleichermaßen beansprucht werden die analogen cDNA-Sequenzen,

insbesondere die volle cDNA-Sequenz des Conductins (Basenpaare 1-3197) gemäß Abb. 2 sowie die Teilsequenzen des Conductins der Nukleotidfolge 452-820 (RGS-Genabschnitt), der Nukleotidfolge 1247-1612 (Genabschnitt der ß-Catenin-Bindungsdomäne) und der Nukleotidfolge 2564-2716 (Genabschnitt der Dishevelled Homologie-Region)

Die Erfindung wird durch die folgenden Ausführungsbeispiele näher erläutert.

Conductin wurde durch einen Hefe 2-Hybrid Screen als B-Catenin-Interaktionspartner identifiziert. Die vollständige cDNA-Sequenz wurde daraufhin isoliert und sequenziert. Die abgeleitete Aminosäuresequenz von Conduction ist in Abb. 1 gezeigt, die Nukleotidsequenz in Abb. 2 und die Gegenüberstellung von Aminpsäure- und Nukleotidsequenz in Abb. 3. Conductin besteht aus 839 Aminosäuren und hat ein Molekulargewicht von 92,4 kDa. Durch Sequenzvergleiche wurde im Conductin eine RGS-Domäne (Aminosäuren 78-200) und eine zu dem Protein Dishevelled verwandte Domäne (Aminosäuren 782-832, Dishevelled Homologie-Region) identifiziert (Abb. 1-3). Die B-Catenin-Bindungsdomäne (Aminosäuren 343-464) wurde durch Interaktionsstudien im 2-Hybrid-System entdeckt (Abb. 4). Es zeigte sich, daß diese Domäne ausreichend und notwendig für die Bindung an B-Catenin ist (Abb. 4), wohingegen die RGS- und Dishevelled Homologie-Region nicht beteiligt sind. Die Wechselwirkung von Conductin mit B-Catenin wurde auch in Co-Immunpräzipitationsexperimenten biochemisch bewiesen.

Die Wirkung von Conductin auf B-Catenin wurde in SW480 Zellen untersucht. In diesen Zellen ist das Tumor-Suppressor-Genprodukt APC mutiert, wodurch es zu einem Anstieg des cytoplasmatischen und vor allem nukleären Gehalts von B-Catenin kommt. Die Einbringung von Conductin in diese Zellen führt zu einem drastischen Abbau von B-Catenin, wodurch die Zelle von cytoplasmatischem und im Zellkern befindlichen B-Catenin depletiert wird (Abb. 4). Diese Wirkung auf den Gehalt B-Catenin

X

ist gleich stark wie die von nichtmutiertem APC, woraus geschlossen werden kann, daß Conductin ebenfalls als Tumorsuppressor durch Regulation von ß-Catenin wirkt. Es wurde außerdem gezeigt, daß Conductin den Wnt/Wingless-Signalweg auch in Xenopus-Embryonen durch seine Wirkung auf ß-Catenin hemmt.

Es wurde außerdem festgestellt, daß Conductin mit APC direkt interagiert. APC-Fragmente von Aminosäure 1464-1604, 1516-1595, 1690-1778 und 1995-2883 wurden als Interaktionsstellen für Conductin identifiziert. In Conductin erfolgt die Bindung an APC über die RGS-Domäne; dieser Bereich ist ausreichend und notwendig für die Interaktion. Die anderen Domänen in Conductin sind nicht beteiligt (Abb. 5).

Legende zu den Abbildungen:

Abb. 1 Aminosauresequenz von Conductin

Die Conductin cDNA kodiert ein Protein von 839 Aminosäuren mit einem berechneten Molekulargewicht von 92,4 kDa. Die RGS-Domäne (doppelt unterstrichen), die B-Catenin-Bindungsdomäne (einfach unterstrichen) und die Dishevelled Homologie-Region sind durch Fettdruck hervorgehoben.

Abb. 2 Nukleotidsequenz von Conductin von Position 1-3199

Die Sequenzbereiche sind analog zu Abb.1 markiert.

Abb. 3 Gegenüberstellung von Aminosäure- und Nukleotidseguenz von Conductin

Abb. 4 Analyse der Interaktion von Conductin und seinen Teilen mit 8-Catenin

Das Conductin Protein und abgeleitete Teilstücke sind schematisch dargestellt. Hervorgehoben sind die RGS-Domäne (RGS) und die B-Catenin-Bindungsstelle (B-BD). Die Interaktion mit B-Catenin wurde im Hefe 2-Hybrid Assay untersucht und als B-Galaktosidase Einheiten quantifiziert. Man erkennt, daß die Bindung an B-Catenin auf die B-Catenin-Bindungsstelle beschränkt ist, die anderen Teile des Proteins tragen dazu nicht bei. Der Abbau von B-Catenin in SW480 Zellen durch Conductin wurde nach transienter Expression der angegeben Proteine und Immunfluoreszenz-Färbung von B-Catenin analysiert. Nur Teilstücke von Conductin, die an B-Catenin binden, führen zu dessen Abbau.

Abb. 5 Analyse der Bindung von Conductin und seinen Teilen an APC

Die Bindung der APC Fragmente von Aminosäure 1464-1604 (APCfr.1)

und 1516-1595 (APCfr. 2) an Conductin und abgeleitete Teilstücke wurden im Hefe 2-Hybrid Assay untersucht und als B-Galaktosidase Einheiten quantifiziert. Die Analyse zeigt die ausschließliche Interaktion von APC mit der RGS-Domäne von Conductin. Vergleichbare Ergebnisse für die Bindung an die RGS-Domäne

wurden mit APC Fragmenten von Aminosäure 1690-1778 und 1995-2883

erhalten.

MSSAVLVTLLPDPSSSFREDAPRPPVPGEEGETPPCQPSVGKVQSTKPMPVSSNARRNED 60
GLGEPEGRASPDSPLTRWTKSLHSLLGDQDGAYLFRTFLEREKCVDTLDFWFACNGFRQM 120
NLKDTKTLRVAKAIYKRYIENNSVVSKQLKPATKTYIRDGIKKQQIGSVMFDQAQTEIQA 180
VMEENAYQVFLTSDIYLEYVRSGGENTAYMSNGGLGSLKVLCGYLPTLNEEEEWTCADLK 240
CKLSPTVVGLSSKTLRATASVRSTETAENGFRSFKRSDPVNPYHVGSGYVFAPATSANDS 300
ELSSDALTDDSMSMTDSSVDGVPPYRMGSKKQLQREMHRSVKANGQVSLPHFPRTHRLPK 360
EMTPVEPAAFAAELISRLEKLKLELESRHSLEERLQQIREDEEKEGSEQALSSRDGAPVQ 420
HPLALLPPAAMKRTHKPFWTTTSPGSSRPPAVNPLVWVAIAHGPAPPTTTTSTTTISSVI 480
PFFRLGASCPVAACPLLGGKSFLTKQTTKHVHHHYIHHHAVPKTKEEIEAEATQRVRCLC 540
PGGTDYYCYSKCKSHPKAPEPLPGEQFCGSRGGTLPKRNAKGTEPGLALSARDGGMSSAA 600
GGPQLPGEEGDRSQDVWQWMLESERQSKSKPHSAQSIRKSYPLESARAAPGERVSRHHLL 660
GASGHSRSVARAHPFTQDPAMPPLTPPNTLAQLEEACRRLAEVSKPQKQRCCVASQQRDR 720
NHSAAGQAGASPFANPSLAPEDHKEPKKLASVHALQASELVVTYFFCGEEIPYRRMLKAQ 780
SLTLGHFKEQLSKKGNYRYYFKKASDEFACGRVFEEIWDDETVLPMYEGRILGKVERID 839

A66.1

AAATAAGCAGCCGTTCGCGATGGATTTCGGGGCCACCCGGAGGCCGAGGCGTCCGCTCCC CAAAGGAGAGCTTTGCTGTAAAAGAGAGGAGGCTCACATGAGCCCCTGCTGACTTAAGAG 120 180 CAAAACAAAATCCAAACTCAGTGAGACGCTCTCCCTCACCATGAGTAGCGCCGTGTTAGT 240 GACTCTCCTTCCAGATCCCAGCAGCAGCTTCCGCGAGGATGCTCCGCGGCCCCCGGTTCC 300 GGGAGAAGAGGGGAGACCCCACCGTGTCAGCCTAGTGTGGGCAAGGTCCAGTCCACCAA 360 ACCTATGCCCGTTTCCTCTAATGCTAGGCGGAATGAAGATGGACTGGGGGAGCCCGAGGG 420 GCGGGCCTCCCCGATTCCCCTTTGACCAGG<u>TGGACCAAGTCTTTACACTCCTTGTTGG</u> 480 TGACCAGGATGGTGCATACCTCTTCCGGACTTTCCTGGAGAGGGAGAAATGTGTGGATAC 540 <u>GCTGGACTTCTGGTTTGCTTGTAATGGGTTCAGGCAGATGAACCTGAAGGATACCAAAAAC</u> 600 TTTGCGAGTGGCCAAAGCAATCTATAAGAGGTACATTGAGAACAACAGCGTTGTCTCCAA 660 GCAGCTGAAGCCCGCCACCAAGACCTACATACGAGATGGCATCAAGAAGCAACAGATCGG 720 CTCGGTCATGTTTGACCAGGCACAGACCGAGATCCAGGCAGTGATGGAGGAAAATGCCTA 780 CCAGGTGTTCTTGACTTCTGACATTTACCTGGAATATGTGAGGAGTGGGGGGGAAAACAC 840 AGCTTACATGAGTAACGGGGGACTGGGGAGCCTAAAGGTCTTATGTGGCTACCTCCCCAC 900 GGTTGGCTTGTCCAGCAAAACTCTTCGGGCCACCGCGAGTGTGAGATCCACGGAAACAGC TGAAAACGGATTCAGGTCCTTCAAGAGAAGCGACCCAGTCAATCCTTATCACGTAGGTTC 1080 CGGCTATGTCTTTGCACCAGCCACCAGTGCCAACGACAGCGAGTTATCCAGCGACGCACT 1140 GACCGACGATTCCATGTCCATGACGGACAGTAGCGTAGATGGAGTCCCTCCTTACCGCAT 1200 GGGGAGTAAGAACAACTCCAGAGAGAGATGCATCGCAGTGTGAAGGCCAATGGCCAAGT 1260 GTCTCTACCTCATTTTCCGAGAACCCACCGCCTGCCCAAGGAGATGACGCCTGTGGAACC 1320 TGCTGCCTTCGCCGAGCTCATCTCCAGGCTGGAGAAACTGAAACTGGAGCTGGAAAG 1380 CCGCCATAGTCTGGAGGAGCGGCTGCAGCAGATCCGGGAGGATGAAGAAAAGGAGGGGTC 1440 TGAGCAGGCCTGAGCTCACGGGATGGAGCACCGGTCCAGCACCCCCTGGCCCTCCTACC 1500 TCCGGCAGCTATGAAGAGGACCCACAAACCATTTTGGACGACCACCTCTCCAGGGTCCTC 1560 AAGACCCCGGCTGTCAATCCCCTGGTGTGGGTCGCTATAGCCCACGGTCCCGCTCCCCC GACCACCACCACCACCACCACCATCAGCAGTGTCATACCCTTCTTTCGACTGGGGGC AAGCTGCCCGTGGCTGCTTGCCCCCTCCTTGGAGGCAAGAGCTTCCTGACCAAACAGAC 1740 GACGAAGCACGTTCACCACCACTACATCCACCACCACGCCGTCCCCAAGACCAAGGAGGA 1800 GATCGAGGCAGAAGCCACACAGAGAGTCCGCTGCCTCTGTCCTGGGGGAACAGATTATTA 1860 TTGTGGCAGCAGAGGTGGTACCTTGCCAAAACGGAATGCAAAGGGCACCGAACCGGGTCT 1980 TGCACTGTCGGCCAGGGATGGAGGGATGTCCAGTGCAGCGGGGGGCCCCCAGCTTCCTGG 2040 GGAAGAAGGAGACCGGTCACAGGATGTCTGGCAGTGGATGTTAGAGAGTGAGCGGCAGAG 2100 CAAGTCCAAGCCCCATAGTGCCCAAAGCATAAGAAAGAGCTACCCATTGGAGTCTGCCCG 2160 TGCGGCCCAGGAGAACGAGTCAGCCGGCACCATCTGTTGGGGGCCAGCGGACACTCCCG 2220 CTCGGTGGCCCGGGCTCACCCATTTACCCAGGACCCTGCAATGCCTCCCCTTACCCCACC 2280 CAACACTTTGGCACAGCTAGAGGAAGCCTGCCGCAGGCTGGCAGAGGTGTCGAAGCCCCA 2340 GAAGCAGCGGTGCTGCGTGGCCAGTCAGCAGAGGGACAGGAACCACTCGGCTGCTGGTCA 2400 GGCAGGAGCCTCACCCTTCGCCAACCCAAGCCTGGCTCCAGAAGATCACAAAGAGCCAAA 2460 GAAACTGGCAAGTGTCCACGCGCTCCAGGCCAGTGAGCTGGTTGTCACCTACTTTTTCTG 2520 TGGAGAAGAATTCCATACAGGAGGATGCTGAAGGCTCAAAGCTTGACCCTGGGCCACTT 2580 CAAGGAGCAGCTCAGCAAAAAGGGAAATTACAGGTATTATTTCAAGAAGGCGAGTGACGA 2640 ATTTGCCTGCGGACGAGTTTTTGAGGAGATCTGGGACGAGACAGTGCTCCCCATGTA 2700 CGAAGGCAGGATCCTGGGCAAAGTGGAGAGGATCGACTGAGCCTTGGCCTCCTCGGCGTG 2760 CAACCTGGGCAAGCACCTCGGCGTGCACCATGGAGCCGAAGCCCAGAGACCCTGTCTCAG 2820 GCCTACGCAACAGCCACGAAATATTCTGAAGGAAAATGAAACCAATTAAGAAGACAAAGC 2880 CTAGGGAGGGACTGGCCCTGGGCCTTCAGGAGGGCGGGGGTATGTTGATCTTCAGTCTC 2940 GGTTCTGAAATTCATAGACTAAGAGAAAACTGTGTATAGCTTGCCCGCACAGGAGTCCTT 3120 ACTGATATTTATTGAACAGTCGATTCCCCCTACCCGCCTCCCTACCCCCCGCCCCCGAGT 3180



TTATGCTGCTTTAAACC

A66.2

1 5'-AAATAAGCAGCCGTTCGCGATGGATTTCGGGGCCACCCGGAGGCCGAGGCGTCCGCTCCCCAAAGG 66 67 AGAGCTTTGCTGTAAAAGAGAGGGGCTCACATGAGCCCCTGCTGACTTAAGAGAGACCAAGC 129 CGATTGCTGAGGGAACTGGAAGAAGAAAAAGGAGGAGGAGGAGAAAAAAAGCAAAACAAAATC 192 CAAACTCAGTGAGACGCTCTCCCTCACCATG AGT AGC GCC GTG TTA GTG ACT CTC A 248 CTT CCA GAT CCC AGC AGC AGC TTC CGC GAG GAT GCT CCG CGG CCC CCG 295 , P P D P S S S F R E D P P 10 A R 25 296 GTT CCG GGA GAA GAA GGG GAG ACC CCA CCG TGT CAG CCT AGT GTG GGC 343 E E G E Т P С P S v 26 G P 0 G 41 344 AAG GTC CAG TCC ACC AAA CCT ATG CCC GTT TCC TCT AAT GCT AGG CGG 391 K M 57 42 392 AAT GAA GAT GGA CTG GGG GAG CCC GAG GGG CGG GCC TCC CCC GAT TCC 439 G 73 Ε D G L Ε Ρ Ε G R S D CCT TTG ACC AGG TGG ACC AAG TCT TTA CAC TCC TTG TTG GGT GAC CAG 487 ___T 74 R <u>W</u> Н 89 GAT GGT GCA TAC CTC TTC CGG ACT TTC CTG GAG AGG GAG AAA TGT GTG 535 488 L F R T F L E R E 90 __c 105 _A_ GAT ACG CTG GAC TTC TGG TTT GCT TGT AAT GGG TTC AGG CAG ATG AAC 536 583 D T L D F W F A C N G F R _ O M 121 CTG AAG GAT ACC AAA ACT TTG CGA GTG GCC AAA GCA ATC TAT AAG AGG 631 K D T K T L R V A K A I Y K R 122 137 TAC ATT GAG AAC AAC AGC GTT GTC TCC AAG CAG CTG AAG CCC GCC ACC 679 632 v s k _0_ 153 N N S L A 680 AAG ACC TAC ATA CGA GAT GGC ATC AAG AAG CAA CAG ATC GGC TCG GTC 727 169 I R_ _D_ ATG TTT GAC CAG GCA CAG ACC GAG ATC CAG GCA GTG ATG GAG GAA AAT 775 M F D O A O T E I O A V M E 185 GCC TAC CAG GTG TTC TTG ACT TCT GAC ATT TAC CTG GAA TAT GTG AGG 776 823 L 0 S D AGT GGG GGG GAA AAC ACA GCT TAC ATG AGT AAC GGG GGA CTG GGG AGC 871 824 T 202 G G Ε N Α Y M S N G G L G S 217 CTA AAG GTC TTA TGT GGC TAC CTC CCC ACC TTG AAT GAA GAA GAG GAG 919 V G Y L P T L N E E E 233 L С E 218 K TGG ACG TGT GCC GAC CTC AAG TGC AAA CTC TCA CCC ACC GTG GTT GGC 920 967 C K S P T 249 TTG TCC AGC AAA ACT CTT CGG GCC ACC GCG AGT GTG AGA TCC ACG GAA 1015 968 265 250 S S K Т L R Α Т Α s R ACA GCT GAA AAC GGA TTC AGG TCC TTC AAG AGA AGC GAC CCA GTC AAT 1063 1016 E F R S F R 281 266 Α N G CCT TAT CAC GTA GGT TCC GGC TAT GTC TTT GCA CCA GCC ACC AGT GCC 1111 1064 Н V G S G Y ν F Α Ρ т 297 1112 AAC GAC AGC GAG TTA TCC AGC GAC GCA CTG ACC GAC GAT TCC ATG TCC 1159 298 D S Ε L S S D Α L Т D D S S 313 ATG ACG GAC AGT AGC GTA GAT GGA GTC CCT CCT TAC CGC ATG GGG AGT 1207 1160 314 M · T D S S v D G v P Р Y S 329 AAG AAA CAA CTC CAG AGA GAG ATG CAT CGC AGT GTG AAG GCC AAT GGC 1255 1208 330 R Ε M H R S v K A 345 1303 1256 CAA GTG TCT CTA CCT CAT TTT CCG AGA ACC CAC CGC CTG CCC AAG GAG Ρ R Т Н 361 F R P 346 н L K ATG ACG CCT GTG GAA CCT GCT GCC TTC GCC GCC GAG CTC ATC TCC AGG 1351 1304 362 E P A A F _A_ A. Ε L s 377 M CTG GAG AAA CTG AAA CTG GAG CTG GAA AGC CGC CAT AGT CTG GAG GAG 1399 1352 393 K Ε L E SRHSL Patt V CGG CTG CAG CAG ATC CGG GAG GAT GAA GAA AAG GAG GGG TCT GAG CAG 1447 1400 L Q Q I R E DEEK E G S E Q 409 394 GCC CTG AGC TCA CGG GAT GGA GCA CCG GTC CAG CAC CCC CTG GCC CTC 1495 1448 S R D G A P V Q Н 425 L S Α

CTA CCT CCG GCA GCT ATG AAG AGG ACC CAC AAA CCA TTT TGG ACG ACC

A A M K R T H K P F

1543

441

W

A66.3

1496

426 L P P

ACC TCT CCA GGG TCC TCA AGA CCC CCG GCT GTC AAT CCC CTG GTG TGG 1591 1544 S Ρ G S S R P P Α ν N 457 1639 1592 GTC GCT ATA GCC CAC GGT CCC GCT CCC CCG ACC ACC ACC ACC ACC ACC 473 <u>P</u> A P P т Т Т T S 458 Α Т Α н G 1687 ACC ACC ATC AGC AGT GTC ATA CCC TTC TTT CGA CTG GGG GCA AGC TGC 1640 S s ν I Ρ F F R G S 489 474 T Ι 1688 CCC GTG GCT GCT TGC CCC CTC CTT GGA GGC AAG AGC TTC CTG ACC AAA 1735 K S Т 505 490 C L G CAG ACG ACG AAG CAC GTT CAC CAC CAC TAC ATC CAC CAC CAC GCC GTC 1783 1736 Н Α 521 506 0 Т Т K н ν Н Н Н Y Ι Н Н CCC AAG ACC AAG GAG GAG ATC GAG GCA GAA GCC ACA CAG AGA GTC CGC 1831 1784 522 Р K T K E Ε Ι E Α Ε Α Т Q R ν 537 TGC CTC TGT CCT GGG GGA ACA GAT TAT TAT TGC 1879 TAC TCC AAA TGC AAA 1832 С S K C K 553 538 D Y AGC CAC CCG AAG GCT CCA GAG CCC CTG CCT GGG GAG CAG TTT TGT GGC 1927 1880 569 C 554 Н Ρ K Α P Ε P L P G E 0 F G AGC AGA GGT GGT ACC TTG CCA AAA CGG AAT GCA AAG GGC ACC GAA CCG 1975 1928 P ĸ R N A K G E 585 570 R G G Т Τ. GGT CTT GCA CTG TCG GCC AGG GAT GGA GGG ATG TCC AGT GCA GCG GGG 1976 2023 586 S Α R D G G М s S G 601 CCT GGG GAA GAA GGA GAC CGG TCA CAG GAT GTC TGG 2071 2024 GGC CCC CAG CTT S 0 D 617 602 G P Q L Ρ G Ε E G D R CAG TGG ATG TTA GAG AGT GAG CGG CAG AGC AAG TCC AAG CCC CAT AGT 2119 2072 W М L Ε s Ε R Q S K S ĸ Ρ Н S 633 618 0 GCC CAA AGC ATA AGA AAG AGC TAC CCA TTG GAG TCT GCC CGT GCG GCC 2167 2120 s Ι R K s Y P Т. E S Α R Α Α 649 634 2215 CCA GGA GAA CGA GTC AGC CGG CAC CAT CTG TTG GGG GCC AGC GGA CAC 2168 G Α G H 665 v Н L L 650 P G E R S R н TCC CGC TCG GTG GCC CGG GCT CAC CCA TTT ACC CAG GAC CCT GCA ATG 2263 2216 Т D Ρ 681 666 ν R Α Н Ρ F 0 Α M CCT CCC CTT ACC CCA CCC AAC ACT TTG GCA CAG CTA GAG GAA GCC TGC 2311 2264 E C 697 А 682 P т P P N Т Τ. Α 0 L E 2359 CGC AGG CTG GCA GAG GTG TCG AAG CCC CAG AAG CAG CGG TGC TGC GTG 2312 V Ρ K Q R 713 Ε S K Q 698 R R L Α GCC AGT CAG CAG AGG GAC AGG AAC CAC TCG GCT GCT CAG GCA GGA 2407 2360 N G Q Α G 729 714 S O Q R D R Н S Α A 2408 GCC TCA CCC TTC GCC AAC CCA AGC CTG GCT CCA GAA GAT CAC AAA GAG 2455 K E 745 E D Н 730 Α S P F Α N P S L Α P CTG GTT 2503 CCA AAG AAA CTG GCA AGT GTC CAC GCG CTC CAG GCC AGT GAG 2456 761 746 Ρ K K L Α s Н Α Q GTC ACC TAC TTT TTC TGT GGA GAA GAA ATT CCA TAC AGG AGG ATG CTG 2551 2504 777 762 F G E Ε I Р Y R R M ν T Y F С 2599 2552 AAG GCT CAA AGC TTG ACC CTG GGC CAC TTC AAG GAG CAG CTC AGC AAA E s ĸ 793 Н F K 0 778 0 S Τ. Т L G AAG GGA AAT TAC AGG TAT TAT TTC AAG AAG GCG AGT GAC GAA TTT GCC 2647 2600 E 809 794 K R Y Y F K K S F Α G TGC GGA CGA GTT TTT GAG GAG ATC TGG GAC GAC GAG ACA GTG CTC CCC 2695 2648 ν F Ē I W D D E T V T. P 825 810 G R E BamHI V ATG TAC GAA GGC AGG ATC CTG GGC AAA GTG GAG AGG ATC GACTGAGCCTT 2745 2696 839 826 G K v Ε $R \cdot I$ L 2808 GGCCTCCTCGGCGTGCAACCTGGGCAAGCACCTCGGCGTGCACCATGGAGCCGAAGCCCAGAG ACCCTGTCTCAGGCCTACGCAACAGCCACGAAATATTCTGAAGGAAAATGAAACCAATTAAGA 2871 AGACAAAGCCTAGGGAGGGACTGGCGCCTGGGCCTTCAGGAGGGCGGGGGTATGTTGATCTTC 2934 2872 2997 2935 3060 2998 GGTTCTGAAATTCATAGACTAAGAGAAAACTGTGTATAGCTTGCCCGCACAGGAGTCCTTACT 3123 3061 3186 3124 3197 3187 TGCTTTAAACC -3'



